

Swiss EMBnet node web server

Laurent Falquet*, Lorenza Bordoli¹, Vassilios Ioannidis, Marco Pagni and C. Victor Jongeneel²

Swiss Institute of Bioinformatics (SIB), CH-1066 Epalinges, Lausanne, Switzerland, ¹Swiss Institute of Bioinformatics (SIB), Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland and ²Swiss Institute of Bioinformatics (SIB) and Office of Information Technology, Ludwig Institute for Cancer Research, CH-1066 Epalinges, Lausanne, Switzerland

Received February 17, 2003; Revised and Accepted March 25, 2003

ABSTRACT

EMBnet is a consortium of collaborating bioinformatics groups located mainly within Europe (<http://www.embnet.org>). Each member country is represented by a 'node', a group responsible for the maintenance of local services for their users (e.g. education, training, software, database distribution, technical support, helpdesk). Among these services a web portal with links and access to locally developed and maintained software is essential and different for each node. Our web portal targets biomedical scientists in Switzerland and elsewhere, offering them access to a collection of important sequence analysis tools mirrored from other sites or developed locally. We describe here the Swiss EMBnet node web site (<http://www.ch.embnet.org>), which presents a number of original services not available anywhere else.

BACKGROUND

The EMBnet organization, founded in 1988, regroups 32 country nodes and seven special nodes in 2003. EMBnet created the first gopher and World Wide Web servers in biology (CSC BioBox <http://www.csc.fi/molbio>). EMBnet developed, among others, solutions for daily database updates using the internet (NDT), distributed computing (HASSLE) and efficient database browsing and linking (SRS) (1). EMBnet is committed to bringing the latest software algorithms to users worldwide free of charge and continues to develop state-of-the-art public domain software (EMBOSS) (2). Many bioinformatics courses at different levels and covering multiple topics are organized every year by EMBnet nodes in various countries.

Altogether, the nodes reach a user base of tens of thousands of scientists around the world and provide them with mirrors of up-to-date databases (e.g. EMBL, SWISS-PROT, TrEMBL, PROSITE, InterPro, ENSEMBL) and software, among which many are only accessible on a particular node web site.

THE Swiss EMBnet WEB SITE

The Swiss EMBnet node was created by Reinhard Dölz in 1988 at the Biozentrum in Basel and in 1997 moved to the SIB (Swiss Institute of Bioinformatics, <http://www.isb-sib.ch>) under the responsibility of Victor Jongeneel and Amos Bairoch. The Swiss EMBnet node currently employs three persons (two full-time positions) funded by a grant from the Swiss government. The services offered through the Swiss EMBnet web site are divided into several categories, e.g. sequence searches in databases, protein function prediction, dotplots, pairwise and multiple alignments, coding region prediction, and tools for the exploration of protein motifs and domains. The Swiss EMBnet node server receives >2300 daily requests for pages, from *.com (18%), *.edu (13%), *.net (7.5%), *.uk (5%), *.ch (4.5%), *.de (4%), *.org (3%), *.fr (2.5%), other (20%), unresolved (22%). User feedback has generally been very positive and the staff aim to respond to suggestions and requests for services or documents.

CONTENT OF THE CURRENT WEB SITE

The latest update of the web site (January 2003) provides the following software tools: (an asterisk indicates that the Swiss EMBnet node web server is an exclusive repository site for the software or provides a unique implementation of the software).

Computer programs

**BLAST (basic and advanced).* Local implementation of the NCBI BLAST2 with many options and databases not seen on other BLAST servers. The output is slightly reformatted to add links to external databases and to display a graphical view of the matches based on a Perl script kindly provided by Dr Alessandro Guffanti (3).

FDF. An access to our local GeneMatcher™ (<http://www.paracel.com>) for fast protein sequence searches using a hardware-implemented version of the Smith–Waterman algorithm. The input forms and output parser are almost identical to those used on the BLAST pages.

*To whom correspondence should be addressed. Tel: +41 216925954; Fax: +41 216925945; Email: laurent.falquet@isb-sib.ch

PRSS3. To evaluate the statistical significance of a pairwise protein sequence alignment (4–6).

LALIGN. To calculate a local or global pairwise alignment between two sequences (part of the FASTA package) (7).

ClustalW. ClustalW multiple sequence alignment software (8).

T-Coffee. T-Coffee multiple sequence alignment software (9).

BOXSHADE. Multiple alignment coloring and formatting tool for publication.

EMBOSS. The Pise interface to the EMBOSS package (2,10).

***Fetch.** A simple tool to retrieve sequences from locally maintained databases.

***HITS.** A local database and many tools devoted to analyzing protein domains and motifs, developed at SIB (11).

***PFSCAN.** A local tool to scan protein sequences for matches with PROSITE patterns profiles and Pfam HMMs (12).

***PFRAMESCAN.** A local tool to scan short DNA sequences (translated on-the-fly into protein) for possible matches with PROSITE profiles (12), allowing for frame-shift errors in the DNA sequence.

PatternFind. A local pattern search tool, allowing the user to enter his/her own pattern and search protein databases for potential matches (12).

***Dotlet.** A Java applet to visualize and tune dot plots interactively (similar to Dotter) developed at the SIB (13).

***iPCR.** Virtual PCR product extractor developed at the SIB.

***ESTScan2.** A tool to detect coding regions in EST-type cDNA sequences with the ability to correct for frame shifts developed at SIB (14).

***TMPRED.** A local tool to detect trans-membrane regions and their orientation (15).

***COILS.** A local tool to detect possible coil-coiled regions in proteins (16).

SAPS. A local tool to collect statistics about proteins sequences (17).

Databases

EPD. The Eucaryotic Promoter Database developed and maintained at SIB (18).

PROSITE. The PROSITE database developed and maintained at SIB (12).

STACK. A repository mirror of the South African STACK database to increase speed downloads in Europe, requires a registration at SANBI (19).

Other

Links to many other useful sites, ftp server with sources code and executables of many software.

FUTURE DEVELOPMENTS

The Swiss EMBnet node plans several developments of its web site. Among them, the implementation of a web interface for Marcoil, a coiled-coil detection tool using HMMs (20). We also plan to offer a local mirror site of the ENSEMBL project (www.ensembl.org) (21), incorporating, in addition, a different collection of genomes. In order to better answer our users requests, we are planning to install a help desk ticketing system and to develop more documentation pages (e.g. FAQs, user manuals). On a more general basis we are committed to maintain and enhance the current set of installed software tools and databases.

REFERENCES

1. Zdobnov, E.M., Lopez, R., Apweiler, R. and Etzold, T. (2002) The EBI SRS server—new features. *Bioinformatics*, **18**, 1149–1150.
2. Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
3. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
4. Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
5. Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.*, **183**, 63–98.
6. Karlin, S. and Altschul, S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl Acad. Sci. USA*, **87**, 2264–2268.
7. Huang, X. and Miller, W. (1991) A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.*, **12**, 337–357.
8. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
9. Notredame, C., Higgins, D. and Heringa, J. (2000) T-Coffee: a novel method for multiple sequence alignments. *J. Mol. Biol.*, **302**, 205–217.
10. Letondal, C. (2001) A web interface generator for molecular biology programs in Unix. *Bioinformatics*, **17**, 73–82.
11. Pagni, M., Iseli, C., Junier, T., Falquet, L., Jongeneel, V. and Bucher, P. (2001) trEST, trGEN and Hits: access to databases of predicted protein sequences. *Nucleic Acids Res.*, **29**, 148–151.
12. Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.*, **30**, 235–238.
13. Junier, T. and Pagni, M. (2000) Dotlet: diagonal plots in a web browser. *Bioinformatics*, **16**, 178–179.
14. Iseli, C., Jongeneel, C.V. and Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 138–148.
15. Hofmann, K. and Stoffel, W. (1993) TMbase—a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, **374**, 166.
16. Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
17. Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B.E. and Karlin, S. (1992) Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl Acad. Sci. USA*, **89**, 2002–2006.
18. Praz, V., Périer, R.C., Bonnard, C. and Bucher, P. (2002) The Eukaryotic Promoter Database, EPD: new entry types and links to gene expression data. *Nucleic Acids Res.*, **30**, 322–324.
19. Christoffels, A., van Gelder, A., Greyling, G., Miller, R., Hide, T. and Hide, W. (2001) STACK: Sequence Tag Alignment and Consensus Knowledgebase. *Nucleic Acids Res.*, **29**, 234–238.
20. Delorenzi, M. and Speed, T. (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, **18**, 617–625.
21. Clamp, M., Andrews, D., Barker, D., Bevan, P., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V. et al. (2003) Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res.*, **31**, 38–42.